5 Transforming Time Series

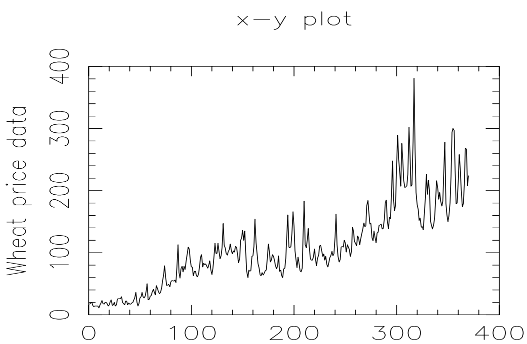
In many situations, it is desirable or necessary to transform a time series data set before using the sophisticated methods we study in this course:

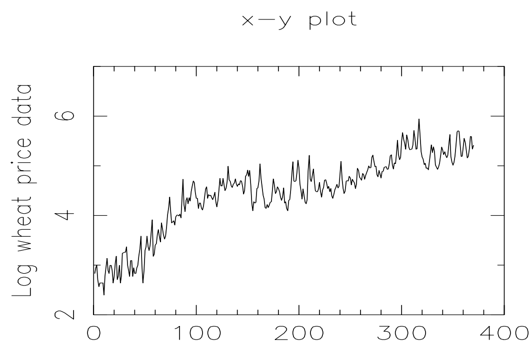
1. Almost all methods assume that the amount of variability in a time series is constant across time.
2. Many times we would like to study what is left in a data set after having removed trends (low frequency content) or cycles in the data.

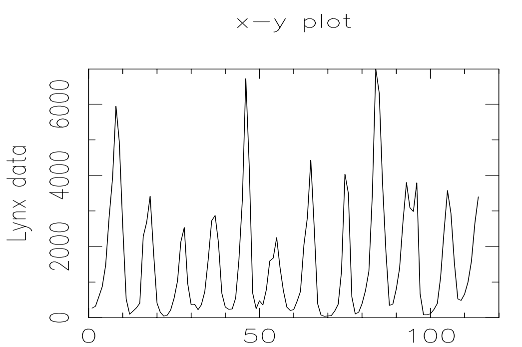
5.1 Power Transformation

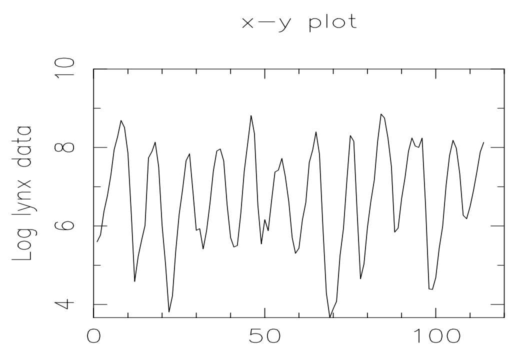
A simple but often effective way to stabilize the variance across time is to apply a power transformation (square root, cube root, log, etc) to the time series.

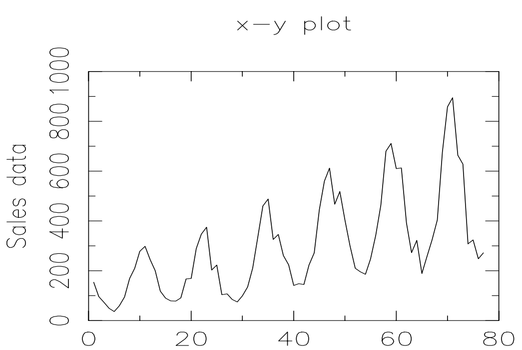
Here are some examples:

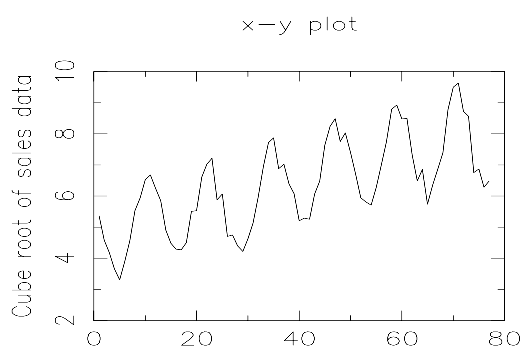












5.2 Dividing Seasonal Standard Deviations

Sometimes with data observed periodically (hourly, daily, monthly, etc), the variability may vary for different periods; for example, there may be more variability on mondays than on tuesdays, and so on. When this happens, it is often useful to calculate the standard deviation for each of the different periods and then for example, divide each monday by the standard deviation of all the mondays, the tuesdays by the standard deviation of the tuesdays, and so on (notice that dividing a set of any numbers by their standard deviation results in the standard deviation of the new set of numbers being equal to one).

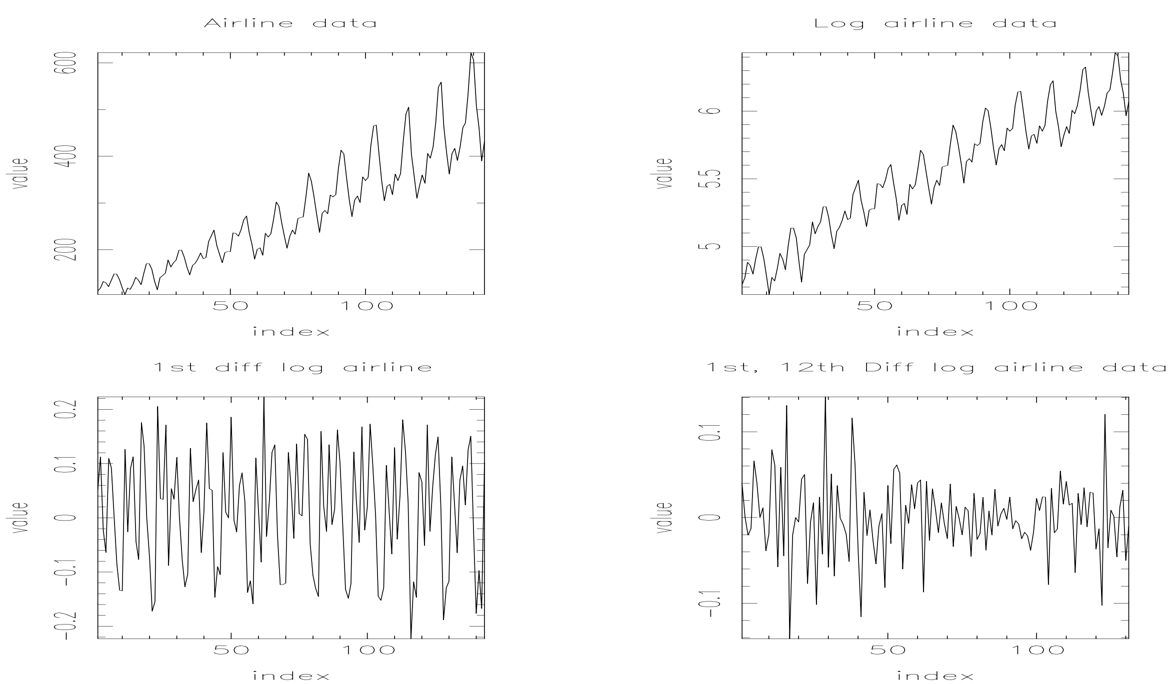
5.3 Subtracting Seasonal Means

One way to remove cycles in data observed periodically is to calculate the sample means of each of the periods (hours or days, for example) and then subtract them from the correpsonding period (subtract the mean of the mondays from mondays, that of tuesdays from each tuesday, and so on).

5.4 Differencing

The dth differencing operator applied to a time series x is to create a new series z whose value at time t is the difference between $x(t+d)$ and $x(t)$. This method works very well in removing trends and cycles. For example, first differencing applied to a series with a linear trend eliminates the trend while if cycles of length d exist in a series, a dth difference will remove them.

Here is a plot of the famous airline data along with its log to stabilize the variance, the 1st difference of the log to eliminate the linear trend, and the 12th difference of the 1st difference of the log to eliminate the annual cycle.



5.5 Regressing on Trends and Cycles

The natural thing for a statistician to do to eliminate trends and cycles in a time series would be to regress $x(t)$ on linear and/or sinusoidal functions of t. For example, we could find the residuals from a model such as

$$x(t) = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi(t-1)/d) + \beta_3 \sin(2\pi(t-1)/d) + \epsilon(t),$$

if we felt there was both a linear trend and a sinusoidal cycle of length d in the data. Note that the X matrix for this regression would be a column of 1's followed by a column $(1, 2, \ldots, n)$, followed by a column of cosines and then a column of sines where the cosines and sines both have amplitude one and period d.

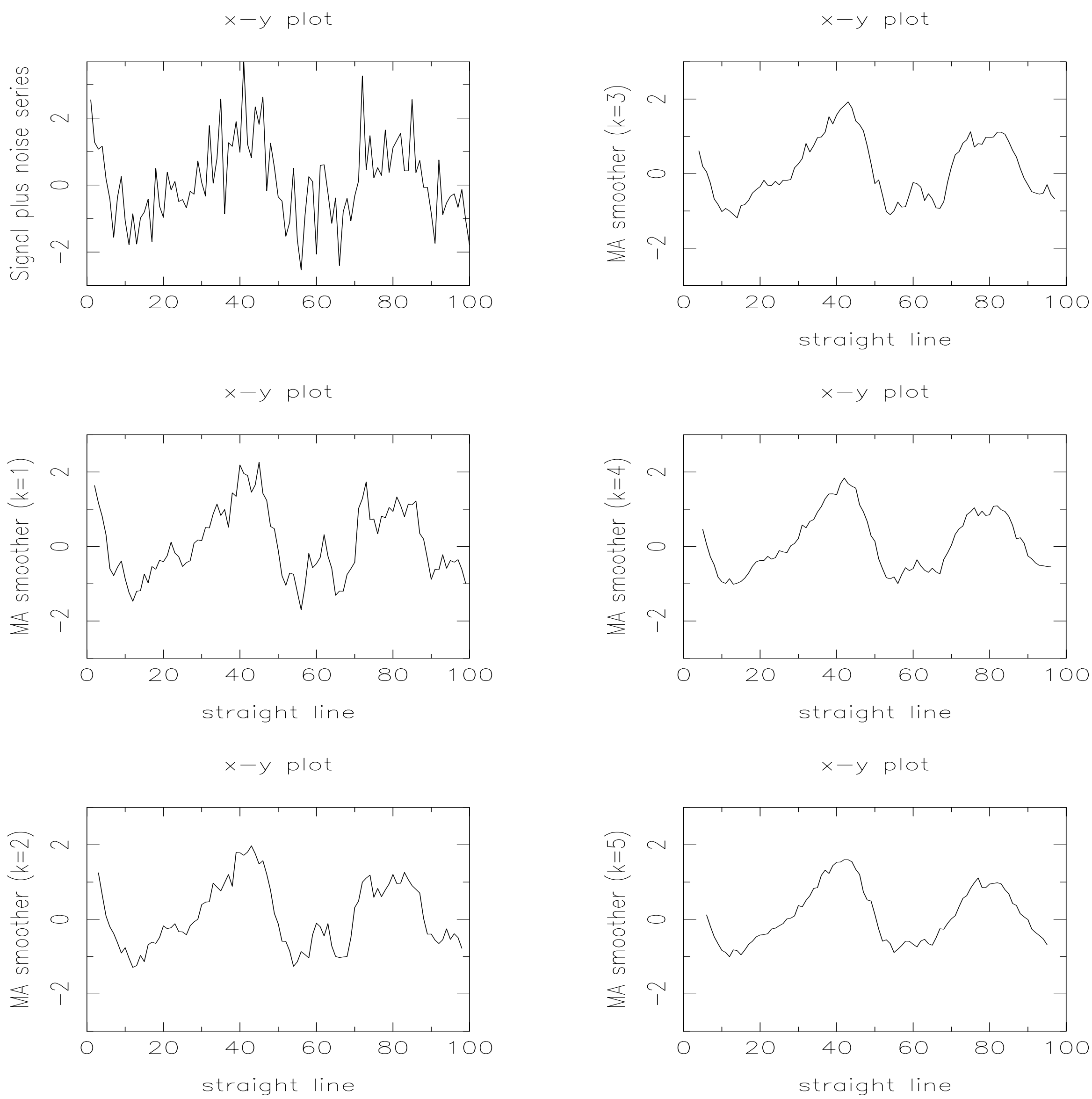
The regression would also give us an idea of the strength of the linear trend from $\hat{\beta}_1$ and/or the sinusoid from $\hat{c} = \sqrt{\hat{\beta}_2^2 + \hat{\beta}_3^2}$.

5.6 Moving Average Smoother

If we produce a new series z whose tth value is the average of $x(t)$ and the K values of x before and the K values after time t, then the result will be smoother than x since consecutive values of z will have many values of x in common in their averages. As K increases, z will get smoother (smaller variance) but values of x further away from time t will be included in $z(t)$ so there will be bias; for example, peaks in x will get chopped off in producing z.

5.7 Example of MA Smoother

Here x is N(0,1) white noise plus a cosine of 100 points with amplitude one and period 40 (noise plus signal). We apply the MA smoother with values of K from one to five. As K increases, the result gets smoother but less representative of original signal.



5.8 General Filters

The MA smoother is a special case of the general idea of using linear smoothers, where new values are weighted averages of old values centered at the time point of interest. An obvious extension would be to use weights other than all $1/(2K+1)$ as in the MA smoother. The weights would be greater for x's near the time point t and smaller farther away from t.

This idea of linear smoother is also called a linear filter.